Variable Frequency Electric Circuit Theory with Application to the Theory of Frequency-Modulation

By JOHN R. CARSON AND THORNTON C. FRY

In this paper the fundamental formulas of variable frequency electric circuit theory are first developed. These are then applied to a study of the transmission, reception and detection of frequency modulated waves. A comparison with amplitude modulation is made and quantitative formulas are developed for comparing the noise-to-signal power ratio in the two modes of modulation.

FREQUENCY modulation was a much talked of subject twenty or more years ago. Most of the interest in it then centered around the idea that it might afford a means of compressing a signal into a narrower frequency band than is required for amplitude modulation. When it was shown that not only could this hope not be realized,* but that much wider bands might be required for frequency modulation, interest in the subject naturally waned. It was revived again when engineers began to explore the possibilities of radio transmission at very short wave lengths where there is little restriction on the width of the frequency band that may be utilized.

During the past eight years a number of papers have been published on frequency modulation, as reference to the attached bibliography will show. That by Professor E. H. Armstrong † deals with this subject in comprehensive fashion. In his paper the problem of discrimination against extraneous noise is discussed, and it is pointed out that important advantages result from a combination of wide frequency bands together with severe amplitude limitation of the received signal waves. His treatment is, however, essentially non-mathematical in character, and it is therefore believed that a mathematical study of this phase of the problem will not be unwelcome. This the present paper aims to supply by developing the basic mathematics of frequency modulation and applying it to the question of noise discrimination with or without amplitude limitation.

The outstanding conclusions reached in the present paper, as regards discrimination against noise by frequency modulation, may be briefly summarized as follows:

* See Bibliography, No. 1.
† See Bibliography, No. 12.

(1) To secure any advantage by frequency modulation as distinguished from amplitude modulation, the frequency band width must be much greater in the former than in the latter system.

(2) Frequency modulation in combination with severe amplitude limitation for the received wave results in substantial reduction of the noise-to-signal power ratio. Formulas are developed which make possible a quantitative estimate of the noise-to-signal power ratio in frequency modulation, with and without amplitude limitation, as compared with amplitude modulation.

It is a pleasure to express our thanks to several colleagues who have been helpful in various ways: to Dr. Ralph Bown who in a brief but very incisive memorandum, which was not intended to be a mathematical study, disclosed all the essential ideas of the quasi-stationary method of attack; to Mr. J. G. Chaffee,* who has been conducting experimental work on frequency modulation in these Laboratories for some years past, by means of which quantitative checks on the accuracy of some of the principal results have been possible; and to various associates, especially Mr. W. R. Bennett and Mrs. S. P. Mead, for detailed criticism of certain portions of the work.

I

In the well-known steady-state theory of alternating currents, the e.m.f. and the currents in all the branches of a network in which the e.m.f. is impressed involve the time t only through the common factor $e^{i\omega t}$ where $i = \sqrt{-1}$ and ω is the *constant* frequency. To this fact is attributable the remarkable simplicity of alternating current theory and calculation, and also the fact that the network is completely specified by its complex admittance $Y(i\omega)$. Thus, if the e.m.f. is $Ee^{i\omega t}$, the steady-state current is

$$I_{ss} = EY(i\omega)e^{i\omega t}. \tag{1}$$

In the present paper we shall deal with the case where the frequency is *variable*, and write the impressed e.m.f. as

$$E \exp\left(i \int_0^t \Omega(t) dt \right). \tag{2}$$

$\Omega(t)$ will be termed the *instantaneous* frequency. This agrees with the usual definition of frequency when Ω is a constant; it is the rate of change of the phase angle at time t; and in addition the interval T between adjacent zeros of $\sin \int \Omega(t)dt$ or $\cos \int \Omega(t)dt$ is approximately $\pi/\Omega(t)$ in cases of practical importance.

* See Bibliography, No. 11.

Instead of dealing with an arbitrary instantaneous frequency $\Omega(t)$ we shall suppose that

$$\Omega(t) = \omega + \mu(t), \tag{3}$$

where ω is a constant and $\mu(t)$ is the variable part of the instantaneous frequency. In practical applications $\mu(t)$ will be written as $\lambda s(t)$ where λ is a real parameter and the mean square value $\overline{s^2}$ of $s(t)$ is taken as equal to 1/2. Other restrictions on $\mu(t)$ will be imposed in the course of the theory to be developed in this paper. Fortunately these restrictions do not interfere with the application of the theory to important problems.

The steady-state current as given by (1) varies with time in precisely the same way as the impressed e.m.f. When the frequency is variable this is no longer true. On the other hand, formula (1) suggests a "quasi-stationary" or "quasi-steady-state current" component, I_{qss}, defined by the formula

$$I_{qss} = EY(i\Omega)\cdot\exp\left(i\int_0^t \Omega dt\right), \tag{4}$$

which corresponds exactly to (1) with the distinction that the admittance is now an explicit function of time. We are thus led to examine the significance of I_{qss} as defined above and the conditions under which it is a valid approximate representation of the actual response of the network to a variable frequency electromotive force, as given by (2).

We start with the fundamental formula of electric circuit theory.[1] Let an e.m.f. $F(t)$ be impressed at time $t = 0$, on a network of indicial admittance $A(t)$; then the current $I(t)$ in the network is given by

$$I(t) = \int_0^t F(t-\tau)\cdot A'(\tau)d\tau. \tag{5}$$

Here $A'(t) = d/dt\cdot A(t)$ and it is supposed that $A(0) = 0$. (This restriction does not limit our subsequent conclusions and is introduced merely to simplify the formulas. Furthermore $A(0)$ is actually zero in all physically realizable networks.)

Omitting the superfluous amplitude constant E we have

$$\begin{aligned} F(t) &= \exp\left(i\int_0^t \Omega dt\right) \\ &= \exp\left(i\omega t + i\int_0^t \mu dt\right), \end{aligned} \tag{6}$$

[1] See J. R. Carson, "Electric Circuit Theory and Operational Calculus," p. 16.

$$\begin{aligned} F(t-\tau) &= \exp\left[i(t-\tau)\omega + i\int_0^{t-\tau} \mu d\tau_1 \right] \\ &= \exp\left[i(t-\tau)\omega + i\int_0^{t} \mu d\tau_1 - i\int_{t-\tau}^{t} \mu d\tau_1 \right] \\ &= \exp\left[i\Omega(t) \right]\cdot \exp\left[-i\omega\tau - i\int_0^{\tau} \mu(t-\tau_1) d\tau_1 \right]. \end{aligned} \tag{7}$$

Substituting this expression in (5) for $F(t-\tau)$ and writing

$$\exp\left(-i\int_0^{\tau} \mu(t-\tau_1)d\tau_1 \right) = M(t, \tau), \tag{8}$$

we have for the current in the network

$$I = e^{i\int \Omega dt}\cdot \int_0^t M(t, \tau)e^{-i\omega\tau}A'(\tau)d\tau. \tag{9}$$

We now split the integral into two parts, thus:

$$\int_0^t = \int_0^\infty - \int_t^\infty .$$

The second integral on the right represents an initial transient which dies away for sufficiently large values of time, t, while the infinite integral represents the total current, I, for sufficiently large values of t. We have therefore

$$\begin{aligned} I &= e^{i\int \Omega dt}\cdot \int_0^\infty M(t, \tau)e^{-i\omega\tau}A'(\tau)d\tau + I_T \\ &= Y(i\omega, t)e^{i\int \Omega dt} + I_T, \end{aligned} \tag{10}$$

where

$$Y(i\omega, t) = \int_0^\infty M(t, \tau)e^{-i\omega\tau}A'(\tau)d\tau. \tag{11}$$

The transient current,[2] I_T, is then given by

$$I_T = e^{i\int \Omega dt} \int_t^\infty M(t, \tau)e^{-i\omega\tau}A'(\tau)d\tau. \tag{12}$$

The foregoing formulas correspond precisely with the formulas for a constant frequency impressed e.m.f.; these are

$$I_{ss} = e^{i\omega t} \int_0^\infty e^{-i\omega\tau}A'(\tau)d\tau, \tag{10a}$$

[2] Hereafter the transient term I_T of (10) will be consistently neglected and the symbol I will refer only to the quasi-stationary current.

$$Y(i\omega) = \int_0^\infty e^{-i\omega\tau} A'(\tau) d\tau, \tag{11a}$$

$$I_T = e^{i\omega t} \int_t^\infty e^{-i\omega\tau} A'(\tau) d\tau, \tag{12a}$$

to which the more general formulas reduce when $\mu = 0$ and consequently $M = 1$.

We have now to evaluate $Y(i\omega, t)$ as given by (11). We shall assume tentatively, at the outset, that $\mu = \lambda s(t)$ has the following properties:

$$\lambda s(t) \ll \omega \quad \text{for all values of } t,$$
$$-1 \leq s(t) \leq 1,$$
$$-1 \leq \int_0^t s dt \leq 1.$$

With these restrictions the instantaneous frequency lies within the limits $\omega \pm \lambda$.

Let us now replace $M(t, \tau)$ by the formal series expansion

$$M(t, \tau) = M(t, 0) + \frac{\tau}{1!}\left[\frac{\partial}{\partial\tau} M(t, \tau)\right]_{\tau=0} + \frac{\tau^2}{2!}\left[\frac{\partial^2}{\partial\tau^2} M(t, \tau)\right]_{\tau=0} + \cdots, \tag{13}$$

which converges in the vicinity of all values of t for which s has a complete set of derivatives. Then, if we write

$$\left[\frac{\partial^n}{\partial\tau^n} M(t, \tau)\right]_{\tau=0} = (-i)^n C_n(t) \tag{13a}$$

and substitute (13) in (11), we get

$$Y(i\omega, t) = \int_0^\infty e^{-i\omega\tau} A'(\tau) d\tau + \sum_1^\infty (-i)^n C_n \int_0^\infty \frac{\tau^n}{n!} e^{-i\omega\tau} A'(\tau) d\tau. \tag{14}$$

From (11a) it follows at once that

$$\int_0^\infty \frac{\tau^n}{n!} e^{-i\omega\tau} A'(\tau) d\tau = \frac{i^n}{n!} \frac{d^n}{d\omega^n} Y(i\omega), \tag{15}$$

so that

$$Y(i\omega, t) = Y(i\omega) + \sum_1^\infty \frac{1}{n!} C_n(t) \frac{d^n}{d\omega^n} Y(i\omega). \tag{16}$$

The coefficients C_n are easily evaluated from (8) and (13*a*); they are [3]

$$\begin{aligned} C_1 &= \mu(t), \\ C_2 &= \mu^2 - i\frac{d}{dt}\mu, \\ &\cdots\cdots\cdots \\ C_{n+1} &= \left(\mu - i\frac{d}{dt}\right)C_n. \end{aligned} \tag{17}$$

Now consider the quasi-stationary admittance $Y(i\Omega)$. Writing $\Omega = \omega + \mu(t)$ and expanding as a power series, we have (assuming that the series is convergent)

$$Y(i\Omega) = Y(i\omega) + \sum_1^\infty \frac{\mu^n}{n!}\frac{d^n}{d\omega^n}Y(i\omega). \tag{18}$$

From (16), (17) and (18) we have at once

$$Y(i\omega, t) = Y(i\Omega) + \sum_2^\infty \frac{1}{n!}D_n(t)\frac{d^n}{d\omega^n}Y(i\omega), \tag{19}$$

where

$$\begin{aligned} D_2 &= -i\frac{d}{dt}\mu(t), \\ D_3 &= -i3\mu\frac{d\mu}{dt} - \frac{d^2}{dt^2}\mu, \\ &\cdots\cdots\cdots \\ D_{m+1} &= C_{m+1} - \mu^{m+1}. \end{aligned} \tag{20}$$

Consequently, the total current, after initial transients have died away, is given by

$$\begin{aligned} I &= I_{qss} + \Delta(t) \\ &= \exp\left(i\int_0^t \Omega dt\right)\cdot\left[Y(i\Omega) - \frac{i}{2!}\frac{d\mu}{dt}\frac{d^2Y}{d\omega^2}\right. \\ &\qquad\left. - \frac{1}{3!}\left(i3\mu\frac{d\mu}{dt} + \frac{d^2\mu}{dt^2}\right)\frac{d^3Y}{d\omega^3} + \cdots\right]. \end{aligned} \tag{21}$$

We have thus succeeded in expressing the response of the network in terms of the quasi-stationary current

$$I_{qss} = Y(i\Omega)\cdot\exp\left(i\int^t \Omega dt\right) \tag{22}$$

[3] From these recursion formulas C_n can be derived in the compact form

$$\begin{aligned} C_n &= \left(\mu - i\frac{d}{dt}\right)\left(\mu - i\frac{d}{dt}\right)\cdots\left(\mu - \frac{d}{dt}\right)\mu \\ &= \left(\mu - i\frac{d}{dt}\right)^{n-1}\mu \quad \textit{symbolically}. \end{aligned}$$

and a correction series Δ, which depends on the derivatives of the steady-state admittance $Y(i\omega)$ with respect to frequency and the derivatives of the variable frequency $\mu(t)$ with respect to time.

If the parameter λ is sufficiently large and the derivatives of s are small enough so that C_n may be replaced by the two leading terms, we get

$$C_n = \mu^n - i\frac{(n-1)n}{2}\mu'\mu^{n-2}, \quad \mu' = \frac{d\mu}{dt}.$$

Then by (16) and (18)

$$\begin{aligned} Y(i\omega, t) &= Y(i\Omega) - \frac{i\mu'}{2}\sum_2^\infty \frac{\mu^{n-2}}{(n-2)!}\frac{d^n}{d\omega^n}Y(i\omega) \\ &= Y(i\Omega) - \frac{i\mu'}{2}\frac{\partial^2}{\partial\mu^2}Y(i\Omega) \\ &= Y(i\Omega) - \frac{i\mu'}{2}\frac{d^2}{d\Omega^2}Y(i\Omega) \\ &= Y(i\Omega) + \frac{i\mu'}{2}Y^{(2)}(i\Omega). \end{aligned} \tag{16a}$$

The preceding formulas are so fundamental to variable frequency theory and the theory of frequency modulation that an alternative derivation seems worth while. We take the applied e.m.f. as

$$E\exp\left(i\omega_c t + i\theta + i\int_0^t \mu dt\right), \tag{23}$$

the phase angle θ being included for the sake of generality.

Now in any finite epoch $0 \leq t \leq T$, it is always possible to write

$$\exp\left(i\int_0^t \mu dt\right) = \int_{-\infty}^{\infty} F(i\omega)e^{i\omega t}d\omega, \tag{24}$$

thus expressing the function on the left as a Fourier integral. For present purposes it is quite unnecessary to evaluate the Fourier function $F(i\omega)$.

Substitution of (24) in (23) gives for the current

$$I = E\cdot\exp\,(i\omega_c t + i\theta)\cdot\int_{-\infty}^{\infty} F(i\omega)Y(i\omega_c + i\omega)e^{i\omega t}d\omega. \tag{25}$$

We suppose as before that, in the interval $0 \leq t \leq T$, $\mu(t)$ and its derivatives are continuous. We can then expand the admittance func-

tion Y in the form

$$
\begin{aligned}
Y(i\omega_c + i\omega) &= Y(i\omega_c) + \frac{i\omega}{1!} Y^{(1)}(i\omega_c) + \frac{(i\omega)^2}{2!} Y^{(2)}(i\omega_c) + \cdots \\
&= Y(i\omega_c) + \sum_1^\infty \frac{(i\omega)^n}{n!} Y^{(n)}(i\omega_c) \\
&= Y(i\omega_c) + \sum_1^\infty \frac{\omega^n}{n!} \frac{d^n}{d\omega_c^n} Y(i\omega_c). \qquad (26)
\end{aligned}
$$

Substitution of (26) in (25) gives

$$
I = E \cdot \exp\,(i\omega_c t + i\theta) \sum_0^\infty \frac{1}{n!} \frac{d^n}{d\omega_c^n} Y(i\omega_c) \int_{-\infty}^{\infty} \omega^n F(i\omega) e^{i\omega t} d\omega. \qquad (27)
$$

But by the identity (24) and repeated differentiations with respect to t, we have

$$
\begin{aligned}
\int_{-\infty}^{\infty} \omega F(i\omega) e^{i\omega t} d\omega &= \mu \exp\left(i \int_0^t \mu dt\right), \\
\int_{-\infty}^{\infty} \omega^2 F(i\omega) e^{i\omega t} d\omega &= \left(\mu^2 - i\frac{d\mu}{dt}\right) \exp\left(i \int_0^t \mu dt\right), \\
&\cdots\cdots\cdots\cdots \qquad (28) \\
\int_{-\infty}^{\infty} \omega^n F(i\omega) e^{i\omega t} d\omega &= C_n \exp\left(i \int_0^t \mu dt\right).
\end{aligned}
$$

Substitution of (28) in (27) gives

$$
I = E \exp\left(i \int_0^t \Omega dt + i\theta\right) \cdot \left\{ Y(i\omega_c) + \sum_1^\infty \frac{1}{n!} C_n \frac{d^n}{d\omega_c^n} Y(i\omega_c) \right\}, \qquad (29)
$$

which agrees with (16).

Formula (25), as it stands, includes the initial transients at time $t = 0$ as well as any which occur at discontinuities in $\mu(t)$. Differentiation with respect to t under the integral sign, however, in effect eliminates these transients and (29) leaves only the quasi-stationary current (plus the correction series given in (19)).

The series appearing in formula (29) may not be convergent; in any case its computation is laborious. Furthermore, in its application to the theory of frequency modulation, terms beyond the first two represent distortion. For these reasons it is often preferable to proceed as follows:

Returning to formula (25), we write

$$
Y(i\omega_c + i\omega) = \left(1 + \frac{\omega}{1!}\frac{d}{d\omega_c} + \cdots + \frac{\omega^n}{n!}\frac{d^n}{d\omega_c^n}\right) Y(i\omega_c) + R_n(\omega_c, \omega), \quad (30)
$$

thus defining the *remainder* R_n. Then (29) becomes

$$I = E \exp\left(i \int_0^t \Omega dt + i\theta\right) \cdot \left[1 + \frac{C_1}{1!}\frac{d}{d\omega_c} + \cdots + \frac{C_n}{n!}\frac{d^n}{d\omega_c{}^n}\right] Y(i\omega_c)$$

$$+ E \exp(i\omega_c t + i\theta) \int_{-\infty}^{\infty} R_n(\omega_c, \omega) F(i\omega) e^{i\omega t} d\omega. \quad (31)$$

In practice it is usually desirable to take $n = 1$.

Now the infinite integral

$$D(t) = \int_{-\infty}^{\infty} R_n(\omega_c, \omega) F(i\omega) e^{i\omega t} d\omega \quad (32)$$

must be kept small if the finite series in (31) is to be an accurate representation of the current I. While it is not in general computable, we see that, in order to keep it small, $R_n(\omega_c, \omega)$ must be small over the essential range of frequencies of $F(i\omega)$. In cases of practical importance we shall find (see Appendix 1) this range is from $\omega = -\lambda$ to $\omega = +\lambda$.

If the transducer introduces a large phase shift, the linear part of which is predominant in the neighborhood of $\omega = \omega_c$, it is preferable to express the received current I in terms of a "retarded" time. To do this, return to (25) and write

$$\begin{aligned} Y(i\omega_c + i\omega) &= |Y(i\omega_c + i\omega)| e^{-i\phi}, \quad (33) \\ \phi &= \omega_c\tau + \omega\tau + \beta(\omega) + \theta_c, \\ \beta(0) &= \beta'(0) = 0, \end{aligned}$$

so that

$$I = E \exp(i\omega_c t' + i\theta') \int_{-\infty}^{\infty} |Y(i\omega_c + i\omega)| e^{-i\beta(\omega)} F(i\omega) e^{i\omega t'} d\omega, \quad (34)$$

where $t' = t - \tau$ is the "retarded" time and $\theta' = \theta - \theta_c$. Formula (34) is identical with (25) but is expressed in the "retarded" time.

Now we can expand the function

$$|Y(i\omega_c + i\omega)| e^{-i\beta(\omega)}$$

in powers of ω; thus

$$\left(1 + \omega \frac{d}{d\omega_c}\right) |Y(i\omega_c)| + \sum_2^{\infty} r_n(\omega_c)\omega^n,$$

where

$$r_n(\omega_c) = \frac{1}{n!}\left\{\frac{\partial^n}{\partial\omega_c{}^n} |Y(i\omega_c + i\omega)| e^{-i\beta(\omega)}\right\}_{\omega=0};$$

and by substitution in (34) we get

$$I = \dot{E} \exp\left(i \int_0^{t'} \Omega(\tau) d\tau + i\theta'\right)$$
$$\times \left[\left(1 + \lambda s(t') \frac{d}{d\omega_c}\right) | Y(i\omega_c) | + \sum_2^{\infty} \frac{r_n}{n!} C_n(t') \right], \quad (35)$$

which corresponds precisely with (29) except that it is expressed in terms of the retarded time t'. If the transducer introduces a large phase delay, (35) may be much more rapidly convergent than (29) and should be employed in preference thereto.

Corresponding to (30) we may write

$$Y(i\omega_c + i\omega)e^{-i\beta(\omega)} = \left(1 + \omega \frac{d}{d\omega_c}\right) | Y(i\omega_c) | + R.$$

which defines the remainder. Then

$$I = E \exp\left(i \int_0^{t'} \Omega d\tau + i\theta'\right) \cdot \left[| Y(i\omega_c) | + \lambda s(t') \frac{d}{d\omega_c} | Y(i\omega_c) | \right]$$
$$+ E \exp (i\omega_c t' + i\theta') D(t'), \quad (36)$$

where

$$D(t') = \int_{-\infty}^{\infty} R(\omega_c, \omega) \cdot F(i\omega) e^{-i\omega t'} d\omega. \quad (37)$$

Formulas (36) and (37) correspond precisely with (31) and (32) and the same remarks apply.

II

The foregoing will now be applied to the Theory of Frequency Modulation. A pure frequency modulated wave may be defined as a high frequency wave of constant amplitude, the "instantaneous" frequency of which is varied in accordance with a low frequency signal wave. Thus

$$W = \exp i \left(\omega_c t + \lambda \int_0^t s(t) dt \right) \quad (38)$$

is a pure frequency modulated wave. Here ω_c is the constant carrier frequency and $s(t)$ is the low frequency signal which it is desired to transmit. λ is a real parameter which will be termed the modulation index. The "instantaneous" frequency is then defined as

$$\omega_c + \lambda s(t).$$

It is convenient to suppose that $s(t)$ varies between ± 1; in this case

the instantaneous frequency varies between the limits

$$\omega_c \pm \lambda.$$

In all cases it will be postulated that $\lambda \ll \omega_c$.

With the method of producing the frequency modulated wave (38) we are not here concerned beyond stating that it may be gotten by varying the capacity or inductance of a high frequency oscillating circuit by and in accordance with the signal $s(t)$.

Corresponding to (38), the pure *amplitude* modulated wave (carrier suppressed) is of the form

$$s(t) \cdot e^{i\omega_c t}. \tag{39}$$

If the maximum essential frequency in the signal $s(t)$ is ω_a, the wave (39) occupies the frequency band lying between $\omega_c - \omega_a$ and $\omega_c + \omega_a$, so that the band width is $2\omega_a$. In the pure *frequency* modulated wave the "instantaneous" frequency band width is 2λ. In practical applications $\lambda \gg \omega_a$. We shall now examine in more detail the concept of "instantaneous" frequency and the conditions under which it has physical significance.

The instantaneous frequency is, as stated, $\omega_c + \lambda s(t)$; a steady-state analysis is of interest and importance. To this end we suppose $s(t) = \cos \omega t$ so that ω is the frequency of the signal. Then the wave (38) may be written

$$e^{i\omega_c t}\left\{\cos\left(\frac{\lambda}{\omega}\sin \omega t\right) + i \sin\left(\frac{\lambda}{\omega}\sin \omega t\right)\right\},$$

and, from known expansions,

$$W = \sum_{n=-\infty}^{\infty} J_n(\lambda/\omega) e^{i(\omega_c + n\omega)t}, \tag{40}$$

where J_n is the Bessel function of the first kind. Thus the frequency modulated wave is made up of sinusoidal components of frequencies

$$\omega_c \pm n\omega, \qquad n = 0, 1, 2, \cdots, \infty.$$

If $\lambda/\omega \gg 1$ (the case in which we shall be interested in practice) the terms in the series (40) beyond $n = \lambda/\omega$ are negligible; this follows from known properties of the Bessel functions. In this case the frequencies lie in the range

$$\omega_c \pm n\omega = \omega_c \pm \lambda,$$

which agrees with the result arrived at from the idea of instantaneous frequency. On the other hand, suppose we make λ so small that $\lambda/\omega \ll 1$. Then (40) becomes to a first order

$$e^{i\omega_c t} + \frac{1}{2}\left(\frac{\lambda}{\omega}\right) e^{i(\omega_c+\omega)t} - \frac{1}{2}\left(\frac{\lambda}{\omega}\right) e^{i(\omega_c-\omega)t},$$

so that the frequencies ω_c, $\omega_c + \omega$, $\omega_c - \omega$ are present in the pure frequency modulated wave.

It is possible to generalize the foregoing and build up a formal steady-state theory by supposing that

$$s(t) = \sum_{m=1}^{M} A_m \cos(\omega_m t + \theta_m). \tag{41}$$

On this assumption, it can be shown that the frequency modulated wave (38) is expressible as

$$W = \exp(i\omega_c t) \prod_m \sum_{n=-\infty}^{\infty} J_n(v_m) \exp[in(\omega_m t + \theta_m)], \tag{42}$$

$$v_m = \lambda A_m/\omega_m.$$

The corresponding current is then

$$\exp(i\omega_c t) \prod_m \sum_{n=-\infty}^{\infty} J_n(v_m) Y(i\omega_c + n\omega_m) \exp[in(\omega_m t + \theta_m)]. \tag{43}$$

Formulas (42) and (43) are purely formal and far too complicated for profitable interpretation. Consequently this line of analysis will not be carried farther.[4]

If we compare the pure *frequency* modulated wave, as given by (38), with the pure *amplitude* modulated wave, as given by (39), it will be observed that, in the latter, the low frequency signal $s(t)$, which is ultimately wanted in the receiver, is *explicit* and methods for its detection and recovery are direct and simple. In the pure frequency modulated wave, on the other hand, the low frequency signal is *implicit;* indeed it may be thought of as concealed in minute phase or frequency variations in the high frequency carrier wave.

If we differentiate (38) with respect to time t, we get

$$dW/dt = [\omega_c + \lambda s(t)] \exp\left(i\omega_c t + i\lambda \int_0^t s dt\right). \tag{44}$$

[4] See Appendix 1.

The first term,

$$\omega_c \exp\left(i\omega_c t + i\lambda\int_0^t s dt\right), \tag{45}$$

is still a pure frequency modulated wave. The second term,

$$\lambda s(t)\cdot\exp\left(i\omega_c t + i\lambda\int_0^t s dt\right), \tag{46}$$

is a "hybrid" modulated wave, since it is modulated with respect to both *amplitude* and *frequency*. The important point to observe is that, by differentiation, we have "rendered explicit" the wanted low frequency signal. We infer from this that the detection of a pure frequency modulated wave involves in effect its differentiation. The process of rendering explicit the low frequency signal has been termed "frequency detection." Actually it converts the *pure frequency* modulated wave into a *hybrid* modulated wave.

Every frequency distorting transducer inherently introduces frequency detection or "hybridization" of the pure frequency-modulated wave, as may be seen from formula (16). The transmitted current is conveniently written in the form

$$I = Y(i\omega_c)\exp\left(i\int_0^t \Omega dt\right)\cdot\left\{1 + \frac{1}{\omega_1}\lambda s + \frac{1}{2!}\frac{1}{\omega_2{}^2}C_2 + \frac{1}{3!}\frac{1}{\omega_3{}^3}C_3 + \cdots\right\}, \tag{47}$$

where

$$\frac{1}{\omega_n{}^n} = \frac{1}{Y(i\omega_c)}\frac{d^n}{d\omega_c{}^n}Y(i\omega_c). \tag{48}$$

(Note that ω_n has the dimensions of frequency. It may be and usually is complex.)

Every term in (47) except the first, is a hybrid modulated wave.

In passing it is interesting to compare the distortion, as given by (47), undergone by the pure *frequency*-modulated wave, with that suffered by the pure *amplitude*-modulated wave (39), in passing through the same transducer. The transmitted current corresponding to the amplitude-modulated wave (39) is

$$I = Y(i\omega_c)e^{i\omega_c t}\left\{s(t) + \frac{1}{i\omega_1}\frac{ds}{dt} + \frac{1}{2!(i\omega_2)^2}\frac{d^2s}{dt^2} + \frac{1}{3!(i\omega_3)^3}\frac{d^3s}{dt^3} + \cdots\right\}. \tag{49}$$

This equation corresponds to (47) for the pure frequency-modulated wave.

III

In this section we consider the recovery of the wanted low frequency signal $s(t)$ from the frequency-modulated wave. This involves two distinct processes: (1) rendering explicit the low frequency signal "implicit" in the high frequency wave; that is, "frequency detection" or "hybridization" of the high frequency wave; and (2) detection proper.

It is convenient and involves no loss of essential generality to suppose that the transducer proper is equalized in the neighborhood of the carrier frequency ω_c; that is,

$$\frac{d}{d\omega_c} Y(i\omega_c), \qquad \frac{d^2}{d\omega_c^2} Y(i\omega_c), \cdots \tag{50}$$

are negligible.

Frequency detection is then effected by a terminal network. We therefore take as the over-all transfer admittance

$$y(i\omega) \cdot Y(i\omega). \tag{51}$$

$y(i\omega)$ represents the terminal receiving network; it is under control and can be designed for the most efficient performance of its function. As we shall see, it should approximate as closely as possible a pure reactance.

Taking the over-all transfer admittance as (51), we have from (47),

$$I = y(i\omega_c) Y(i\omega_c) \cdot \exp\left(i \int_0^t \Omega dt \right) \times \left\{ 1 + \frac{1}{\omega_1} \lambda s + \frac{1}{2!\omega_2^2} C_2 + \frac{1}{3!\omega_3^3} C_3 + \cdots \right\}, \tag{52}$$

where now

$$1/\omega_n{}^n = \frac{1}{y(i\omega_c)} \frac{d^n}{d\omega_c{}^n} y(i\omega_c). \tag{53}$$

Inspection of (52) shows that the terms beyond the second simply represent distortion. The terminal network or frequency detector should be so designed as to make the series

$$1 + \frac{\lambda}{\omega_1} + \left(\frac{\lambda}{\omega_2}\right)^2 + \left(\frac{\lambda}{\omega_3}\right)^3 + \cdots$$

rapidly convergent from the start.[5] In fact the ideal frequency detector is a network whose admittance $y(i\omega)$ can be represented with

[5] See note at end of this section (p. 528) for specific example.

sufficient accuracy in the neighborhood of $\omega = \omega_c$ by the expression

$$y(i\omega) = y(i\omega_c)\left(1 + \frac{\omega - \omega_c}{\omega_1}\right). \tag{53a}$$

This approximation should be valid over the frequency range from $\omega = \omega_c - \lambda$ to $\omega = \omega_c + \lambda$.

Supposing that this condition is satisfied, the wave, after passing over the transducer and through the terminal frequency detector, is (omitting the constant $y \cdot Y$)

$$I = \left(1 + \frac{\lambda}{\omega_1} s(t)\right) \cdot \exp\left(i \int_0^t \Omega dt\right). \tag{54}$$

If y is a pure reactance, ω_1 is a pure real; due to unavoidable dissipation it will actually be complex. To take this into account we replace ω_1 in (54) by $\omega_1 e^{-i\alpha}$ where now ω_1 is real; (54) then becomes

$$I = \left\{1 + \frac{\lambda}{\omega_1} \cos \alpha \cdot s(t) + i \frac{\lambda}{\omega_1} \sin \alpha \cdot s(t)\right\} \exp\left(i \int_0^t \Omega dt\right). \tag{55}$$

The amplitude A of this wave is then

$$A = \left\{\left(1 + \frac{\lambda}{\omega_1} \cos \alpha \cdot s(t)\right)^2 + \left(\frac{\lambda}{\omega_1} \sin \alpha \cdot s(t)\right)^2\right\}^{1/2}. \tag{56}$$

Now let λ/ω_1 be *less than unity* and let the wave (55) be impressed on a straight-line rectifier. Then the rectified or detected output is

$$\left(1 + \frac{\lambda}{\omega_1} \cos \alpha \cdot s(t)\right)\left\{1 + \left(\frac{\lambda \sin \alpha \cdot s(t)}{\omega_1 + \lambda \cos \alpha \cdot s(t)}\right)^2\right\}^{1/2}, \tag{57}$$

or, to a first order,

$$1 + \frac{\lambda}{\omega_1} \cos \alpha \cdot s(t) + \frac{1}{2} \frac{\lambda^2}{\omega_1^2} \sin^2 \alpha \cdot s^2(t). \tag{58}$$

The second term is the recovered signal and the third term is the first order non-linear distortion.

Inspection of the foregoing formulas shows at once that, for detection by straight rectification, the following conditions should be satisfied:

(1) λ/ω_1 *must* be less than unity.
(2) The terminal network should be as nearly as possible a pure reactance to make the phase angle α as nearly zero as possible.

(3) To minimize both linear and non-linear distortion it is necessary that the sequence

$$\frac{\lambda}{\omega_1}, \quad \left(\frac{\lambda}{\omega_2}\right)^2, \quad \left(\frac{\lambda}{\omega_3}\right)^3, \cdots$$

be rapidly convergent from the start.

The first term of (58) is simply direct current and has no significance as regards the recovered signal. When we come to consider the problem of noise in the next section, we shall find that its elimination is important. This can be effected by a scheme which may be termed *balanced rectification*. Briefly described the scheme consists in terminating the transducer in two frequency detectors y_1 and y_2 in parallel; these are so adjusted that $y_1(i\omega_c) = -y_2(i\omega_c)$ and $dy_1/d\omega_c = dy_2/d\omega_c$. ω_1 is therefore of opposite sign in the two frequency detectors. The rectified outputs of the two parallel circuits are then differentially combined in a common low frequency circuit. Corresponding to (58), the resultant detected output is

$$2\frac{\lambda}{\omega_1}\cos\alpha\cdot s(t). \tag{59}$$

This arrangement therefore eliminates first order non-linear distortion, as well as the constant term.

Rectification is the simplest and most direct mode of detection of frequency-modulated waves. However, in connection with the problem of noise reduction other methods of detection will be considered.

Note

As a specific example of the foregoing let the terminal frequency detector, specified by the admittance $y(i\omega)$, be an oscillation circuit consisting simply of an inductance L in series with a capacitance C. Then

$$y(i\omega) = i\sqrt{\frac{C}{L}}\frac{\omega/\omega_R}{1-\omega^2/\omega_R^2},$$

where $\omega_R^2 = 1/LC$.

Then, if ω_c/ω_R is nearly equal to unity, that is, if

$$\omega_R = (1+\delta)\omega_c,$$
$$|\delta| \ll 1,$$

we have approximately,

$$\frac{1}{\omega_n{}^n} \doteq \frac{n!}{(\omega_R-\omega_c)^n},$$

$$y(i\omega_c) \doteq \frac{i}{2}\frac{\sqrt{C/L}}{\omega_R-\omega_c}.$$

Formula (42) thus becomes

$$I = y(i\omega_c) \cdot Y(i\omega_c) \cdot \exp\left(i \int_0^t \Omega dt\right) \cdot \left\{1 + \frac{\lambda s}{\omega_R - \omega_c} + \frac{C_2}{(\omega_R - \omega_c)^2} + \frac{C_3}{(\omega_R - \omega_c)^3} + \cdots\right\}.$$

In order that the distortion shall be small it is necessary that

$$\lambda \ll |\omega_R - \omega_c|.$$

If the two networks y_1 and y_2 are oscillation circuits so adjusted that

$$\begin{aligned} C_1/L_1 &= C_2/L_2, \\ \omega_{R_1} &= (1 + \delta)\omega_c = 1/\sqrt{L_1C_1}, \\ \omega_{R_2} &= (1 - \delta)\omega_c = 1/\sqrt{L_2C_2}, \end{aligned}$$

then the combined rectified output of the two parallel circuits is proportional to

$$\frac{\lambda s}{\delta \cdot \omega_c} + \frac{C_3}{(\delta \cdot \omega_c)^3} + \frac{C_5}{(\delta \cdot \omega_c)^5} + \cdots.$$

Thus the constant term and the first order distortion are eliminated in the low frequency circuit.

IV

The most important advantage known at present of *frequency*-modulation, as compared with *amplitude*-modulation, lies in the possibility of substantial reduction in the low frequency noise-to-signal power ratio in the receiver. Such reduction requires a correspondingly large increase in the width of the high frequency transmission band. For this reason frequency-modulation appears to be inherently restricted to short wave transmission.

In the discussion of the theory of noise which follows, it is expressly assumed *that the high frequency noise is small compared with the high frequency signal wave*. Also ideal terminal networks, filters and detectors are postulated.

In view of the assumption of a low noise power level, the calculation of the low frequency noise power in the receiver proper can be made to depend on the calculation of the noise due to the typical high frequency noise element

$$A_n \exp(i\omega_c t + i\omega_n t + i\theta_n). \tag{60}$$

Corresponding to the noise element (60), the output of the ideal frequency detector is

$$\exp\left(i\int_0^t \Omega dt\right)\cdot\left\{1+\frac{\lambda s}{\omega_1}+\left(1+\frac{\omega_n}{\omega_1}\right)A_n\exp\left(i\omega_n t+i\theta_n - i\lambda\int_0^t s dt\right)\right\}. \quad (61)$$

Since the expression

$$\exp\left(i\omega_n t+i\theta_n - i\lambda\int_0^t s dt\right)$$

occurs so frequently in the analysis which is to follow, it is convenient to adopt the notation

$$\Omega_n = \omega_n - \lambda s(t),$$
$$\int_0^t \Omega_n dt = \omega_n t - \lambda\int_0^t s dt. \quad (61a)$$

With this notation and on the assumption that $A_n \ll 1$ and ω_1 real, the amplitude of the wave (61) is

$$1+\frac{\lambda s}{\omega_1}+\left(1+\frac{\omega_n}{\omega_1}\right)A_n\cos\left(\int_0^t \Omega_n dt\right). \quad (62)$$

In this formula the argument of the cosine function should be strictly

$$\int_0^t \Omega_n dt + \theta_n.$$

The phase angle θ_n is random however and does not affect the final formulas; it may therefore be omitted at the outset. Consequently, if the wave (61) is passed through a straight line rectifier, the rectified or low frequency current is proportional to

$$\lambda s(t) + (\omega_1+\omega_n)A_n\cos\left(\int_0^t \Omega_n dt\right). \quad (63)$$

The first term is the recovered signal and the second term the low frequency noise or interference corresponding to the high frequency element (60).

Now the wave (63), before reaching the receiver proper, is transmitted through a low-pass filter, which cuts off all frequencies above ω_a; ω_a is the highest essential frequency in the signal $s(t)$. Consequently, in order to find the noise actually reaching the receiver proper, it is

necessary in one way or another to make a frequency analysis of the wave (63). This is done in Appendix 2, attached hereto, where however, instead of dealing with the special formula (63), a more general expression

$$\lambda s(t) + (\omega_1 + \omega_n + \mu s)A_n \cos \int_0^t \Omega_n dt, \tag{64}$$

is used for the low frequency current. This will be found to include, as special cases, several other important types of rectification, as well as amplitude limitation, which we shall wish to discuss later.[6] Then, subject to the limitation that the noise energy is uniformly distributed over the spectrum, it is shown in Appendix 2 that

$$P_S = \lambda^2 \overline{s^2}, \tag{65}$$

$$P_N = (\tfrac{1}{3}\omega_a^2 + \omega_1^2 + (1 + \nu)^2 \lambda^2 \overline{s^2})\omega_a N^2, \tag{66}$$

$$\nu = \mu/\lambda, \tag{67}$$

N^2 = mean high frequency power level.

These formulas are quite important because they make the calculation of low frequency noise-to-signal power ratio very simple for all the modes of frequency detection and demodulation which we shall discuss.

Applying them to formula (63) we find for *straight line rectification*

$$P_N = (\tfrac{1}{3}\omega_a^2 + \omega_1^2 + \lambda^2 \overline{s^2})\omega_a N^2, \tag{68}$$
$$P_S = \lambda^2 \overline{s^2}.$$

It is known that in practice $\omega_1^2 \gg \lambda^2 \overline{s^2}$ and $\lambda^2 \overline{s^2} \gg \omega_a^2$. Consequently in the factor $(\tfrac{1}{3}\omega_a^2 + \omega_1^2 + \lambda^2 \overline{s^2})$ the largest term is ω_1^2. Therefore it is important, if possible, to eliminate this term. This can be effected by the scheme briefly discussed at the close of section III; parallel rectification and differential recombination. For this scheme the low frequency current is found to be proportional to

$$\lambda s + \omega_n A_n \cos\left(\int_0^t \Omega_n dt\right). \tag{69}$$

Consequently, for *parallel rectification* and *differential recombination*,

$$P_N = (\tfrac{1}{3}\omega_a^2 + \lambda^2 \overline{s^2})\omega_a N^2. \tag{70}$$

[6] The formula is also general enough to include detection by a product modulator, which however is not discussed in the text as no advantage over linear rectification was found.

Here, in the factor $(\frac{1}{3}\omega_a^2 + \lambda^2\overline{s^2})$, the term $\lambda^2\overline{s^2}$ is predominant. The elimination of the term ω_1^2 has resulted in a substantial reduction in the noise power.

Returning to the general formula (66) for P_N, it is clear, that, if in addition to eliminating the term ω_1^2, the parameter $\nu = \mu/\lambda$ can be made equal to -1, the noise power will be reduced to its lowest limits:

$$P_N = \tfrac{1}{3}\omega_a^3 N^2.$$

This highly desirable result can be effected by *amplitude limitation*, the theory of which will now be discussed.

V

When amplitude limitation is employed in frequency-modulation, the incoming high frequency signal is drastically reduced in amplitude. If no interference is present this merely results in an equal reduction in the low frequency recovered signal which is *per se* undesirable. When, however, noise or interference is present, amplitude limitation prevents the interference from affecting the *amplitude* of the resultant high frequency wave; its effect then can appear only as *variations in the phase or instantaneous frequency* of the high frequency wave. To this fact is to be ascribed the potential superiority of *frequency*-modulation as regards the reduction of noise power. This superiority is only possible with wide band high frequency transmission; that is, the index of frequency-modulation λ must be large compared with the low frequency band width ω_a. Insofar as the present paper is concerned, the potential superiority of frequency-modulation with amplitude limitation is demonstrated only for the case where the high frequency noise is small compared with the high frequency signal wave.

If, to the frequency-modulated wave $\exp\left(i\int_0^t \Omega dt\right)$, there is added the typical noise element $A_n \exp(i\omega_c + i\omega_n t + \theta_n)$, the resultant wave may be written as

$$\exp\left(i\int_0^t \Omega dt\right)\cdot\left(1 + A_n \exp\left(i\int_0^t \Omega_n dt\right)\right). \tag{71}$$

Postulating that $A_n \ll 1$ and therefore neglecting terms in A_n^2, the real part of (71) is

$$\left(1 + A_n \cos\left(\int_0^t \Omega_n dt\right)\right)\cdot\cos\left(\int_0^t \Omega dt + A_n \sin\left(\int_0^t \Omega_n dt\right)\right). \tag{72}$$

If this wave is subjected to amplitude limitation, the amplitude variation is suppressed, leaving a pure frequency-modulated wave, *proportional* to the *real part* of

$$\exp\left[i \left(\int_0^t \Omega dt + A_n \sin \left(\int_0^t \Omega_n dt \right) \right) \right] \tag{73}$$

(but drastically reduced in amplitude).

After frequency detection the wave (73) is, within a constant,

$$\exp\left[\left(i \int_0^t \Omega dt + A_n \sin \left(\int_0^t \Omega_n dt \right) \right) \right] \times \left[1 + \frac{1}{\omega_1} \frac{d}{dt} \left(\lambda \int_0^t s dt + A_n \sin \left(\int_0^t \Omega_n dt \right) \right) \right]. \tag{74}$$

Consequently, since

$$\int_0^t \Omega_n dt = \omega_n t + \theta_n - \lambda \int_0^t s dt, \tag{75}$$

the amplitude of the wave (74) is

$$1 + \frac{1}{\omega_1} \left\{ \lambda s + (\omega_n - \lambda s) A_n \cos \left(\int_0^t \Omega_n dt \right) \right\}. \tag{76}$$

This is the amplitude of the low frequency wave after rectification; it is obviously proportional to

$$\lambda s + (\omega_n - \lambda s) A_n \cos \left(\int_0^t \Omega_n dt \right), \tag{77}$$

which is a special case of (64) and may be used in calculating the relative signal and noise power with amplitude limitation. Hence we have, by aid of (65) and (66),

$$\begin{aligned} P_S &= \lambda^2 \overline{s^2}, \\ P_N &= \tfrac{1}{3} \omega_a{}^3 N^2. \end{aligned} \tag{78}$$

(These are, of course, relative values and take no account of the absolute reduction in power due to amplitude limitation.)

Comparing (78) with (68) it is seen that, for detection by straight line rectification, the ratio of the noise power *with* to that *without* amplitude limitation is

$$\frac{1}{1 + 3\omega_1{}^2/\omega_a{}^2 + 3\lambda^2 \overline{s^2}/\omega_a{}^2}; \tag{79}$$

or taking $\overline{s^2} = 1/2$,

$$\frac{1}{1 + 3\omega_1^2/\omega_a^2 + 3\lambda^2/2\omega_a^2}. \tag{80}$$

Since in practice $\omega_1 \gg \omega_a$ and $\lambda \gg \omega_a$, amplitude limitation results in a very substantial reduction in low frequency noise power in the receiver proper. Reference to formula (70) shows that, as compared with parallel rectification and recombination, amplitude limitation reduces the noise power by the factor

$$\frac{1}{1 + 3\lambda^2/2\omega_a^2}. \tag{81}$$

It should be observed that *without* amplitude limitation little reduction in the noise-to-signal power ratio results from increasing the modulation index λ (and consequently the high frequency transmission band width). On the other hand, *with* amplitude limitation, the ratio ρ of noise-to-signal power is

$$\rho = P_N/P_S = \frac{2}{3}\left(\frac{\omega_a}{\lambda}\right)^2 \omega_a N^2. \tag{82}$$

The ratio ρ is then (within limits) inversely proportional to the square of the modulation index λ, so that a large value of λ is indicated. It should be noted that, within limits ($\lambda \ll \omega_c$), the power transmitted from the sending station is independent of the modulation index λ.

It might be inferred from formula (82) that the noise power ratio ρ can be reduced indefinitely by indefinitely increasing the modulation index λ. Actually there are practical limits to the size of λ. First, if λ is made sufficiently large, the variable frequency oscillator generating the frequency-modulated wave may become unstable or function imperfectly. Secondly, the frequency spread of the frequency modulated wave is 2λ (from $\omega_c - \lambda$ to $\omega_c + \lambda$) and, if this is made too large, interference with other stations will result. Finally, the stationary distortion of the recovered low frequency signal $s(t)$ increases rapidly with the size of λ.

To summarize the results of the foregoing analysis the potential advantages of frequency-modulation depend on two facts. (1) By increasing the modulation index λ it is possible to increase the recovered low frequency signal power at the receiving station without increasing the high frequency power transmitted from the sending station. (2) It is possible to employ amplitude limitation (inherently impossible with amplitude-modulation) whereby the effect of interference or noise is reduced to a phase or "instantaneous frequency" variation of the high frequency wave.

Appendix 1

Formula (40) *et sequa* establish the fact that the actual frequency of the wave (29) varies between the limits

$$\omega_c \pm \lambda$$

provided $s(t)$ is a pure sinusoid $\lambda \sin \omega t$ and $\lambda \gg \omega$. This agrees with the concept of instantaneous frequency.

When $s(t)$ is a complex function—say a Fourier series—the frequency range of W can be determined qualitatively under certain restrictions, as follows:

We write

$$W = \exp\left(i\omega_c t + i\lambda \int_0^t s dt\right) \tag{1a}$$

$$= e^{i\omega_c t} \int_{-\infty}^{\infty} F(i\omega) e^{i\omega t} d\omega. \tag{2a}$$

The Fourier formulation is supposed to be valid in the epoch $0 \leq t \leq T$ and T can be made as great as desired. Then

$$F(i\omega) = \pi \int_0^T \exp\left(i\lambda \int_0^t s dt - i\omega t\right) dt. \tag{3a}$$

We now suppose that, in the epoch $0 \leq t \leq T$,

$$\left| \lambda \int_0^T s dt \right| \tag{4a}$$

becomes very large compared with 2π. On this assumption, it follows from the Principle of Stationary Phase, that, for a fixed value of ω, the important contributions to the integral (3a) occur for those values of the integration variable t for which

$$\frac{d}{dt}\left(\lambda \int_0^t s dt - \omega t\right) = 0,$$

or

$$\omega = \lambda s(t).$$

Consequently the important part of the spectrum $F(i\omega)$ corresponds to those values of ω in the range

$$\lambda s_{\min} \leq \omega \leq \lambda s_{\max}.$$

Therefore the frequency spread of W lies in the range from $\omega_c + \lambda s_{\min}$ to $\omega_c + \lambda s_{\max}$ or $\omega_c \pm \lambda$ if $s_{\max} = - s_{\min} = 1$.

Appendix 2

We take the frequency modulated wave as

$$\cos\left(\omega_c t + \lambda \int_0^t s dt\right), \tag{1b}$$

where ω_c is the carrier frequency and $s = s(t)$ is the low frequency signal. λ is a real parameter, which fixes the amplitude of the frequency spread.

Correspondingly, we take the typical noise element as

$$A_n \cos ((\omega_c + \omega_n)t + \theta_n). \tag{2b}$$

For reasons stated in the text, we take the more general formula for the low frequency current as proportional to

$$\lambda s + (\omega_0 + \omega_n + \mu s)A_n \cos\left(\omega_n t + \theta_n - \lambda \int_0^t s dt\right), \tag{3b}$$

where ω_0, λ, μ are real parameters. The term λs is the recovered signal and the second term is the low frequency noise corresponding to the high frequency noise element (2*b*).

We suppose that the noise is uniformly distributed over the frequency spectrum, at least in the neighborhood of $\omega = \omega_c$, so that, corresponding to the noise element

$$A_n \cos (\omega_n t + \theta_n), \tag{4b}$$

the noise is representable as the Fourier integral

$$\frac{N}{\pi} \int \cos (\omega_n t + \theta_n) d\omega_n \tag{5b}$$

and the corresponding *noise power* for the frequency interval $\omega_1 < \omega_n < \omega_2$ is, by the Fourier integral energy theorem,

$$\frac{N^2}{\pi} \int_{\omega_1}^{\omega_2} d\omega_n = \frac{1}{\pi}(\omega_2 - \omega_1)N^2. \tag{6b}$$

The Fourier integral energy theorem states that, if in the epoch $0 \leq t \leq T$, the function $f(t)$ is representable as the Fourier integral

$$f(t) = \frac{1}{\pi} \int_0^\infty F(\omega) \cdot \cos (\omega t + \theta(\omega)) d\omega, \tag{7b}$$

then

$$\int_0^T f^2 dt = \frac{1}{\pi}\int_0^\infty F^2 d\omega.^7 \tag{8b}$$

Replacing (4*b*) by (5*b*) to take care of the distributed noise, the noise term of (3*b*) becomes

$$\cos\left(\lambda\int_0^t s dt\right)\cdot\frac{N}{\pi}\int(\omega_0+\omega_n+\mu s)\cdot\cos(\omega_n t+\theta_n)d\omega_n$$
$$+\sin\left(\lambda\int_0^t s dt\right)\cdot\frac{N}{\pi}\int(\omega_0+\omega_n+\mu s)\cdot\sin(\omega_n t+\theta_n)d\omega_n. \tag{9b}$$

Now this noise in the low frequency circuit is passed through a low pass filter, which cuts off all frequencies above ω_a. ω_a is the maximum essential frequency in the signal $s(t)$.

It is therefore necessary to express (9*b*) as a frequency function before calculating the noise power. To this end we write the Fourier integrals

$$\cos\left(\lambda\int_0^t s dt\right)=\frac{1}{\pi}\int_0^\infty F_c\cos(\omega t+\theta_c)d\omega, \tag{10b}$$

$$\sin\left(\lambda\int_0^t s dt\right)=\frac{1}{\pi}\int_0^\infty F_s\sin(\omega t+\theta_s)d\omega. \tag{11b}$$

We note also that

$$\mu s\cdot\cos\left(\lambda\int_0^t s dt\right)=\frac{\mu}{\lambda}\frac{d}{dt}\sin\left(\lambda\int_0^t s dt\right)$$
$$=\frac{1}{\pi}\int_0^\infty\frac{\mu\omega}{\lambda}F_s\cos(\omega t+\theta_s)d\omega, \tag{12b}$$

$$\mu s\cdot\sin\left(\lambda\int_0^t s dt\right)=-\frac{\mu}{\lambda}\frac{d}{dt}\cos\left(\lambda\int_0^t s dt\right)$$
$$=\frac{1}{\pi}\int_0^\infty\frac{\mu\omega}{\lambda}F_c\sin(\omega t+\theta_c)d\omega. \tag{13b}$$

Substituting (10*b*), (11*b*), (12*b*) and (13*b*) in (9*b*) and carrying through straightforward operations, we find that the noise is given by

$$\frac{N}{2\pi^2}\int_0^\infty F_p d\omega\int_{\omega-\omega_a}^{\omega+\omega_a}\left(\omega_0+\omega_n+\frac{\mu}{\lambda}\omega\right)\cos((\omega-\omega_n)t+\theta_p)d\omega_n$$
$$+\frac{N}{2\pi^2}\int_0^\infty F_m d\omega\int_{-(\omega+\omega_a)}^{-(\omega-\omega_a)}\left(\omega_0+\omega_n-\frac{\mu}{\lambda}\omega\right)\cos((\omega+\omega_n)t+\theta_m)d\omega_n, \tag{14b}$$

[7] See "Transient Oscillations in Electric Wave Filters," Carson and Zobel, *B. S. T. J.*, July, 1923.

where

$$F_p^2 = F_c^2 + F_s^2 + 2F_cF_s \cos(\theta_c - \theta_s), \tag{15b}$$

$$F_m^2 = F_c^2 + F_s^2 - 2F_cF_s \cos(\theta_c - \theta_s). \tag{16b}$$

The limits of integration of ω_n are determined by the fact that, $\omega - \omega_n$ in the first integral of (14*b*) and $\omega + \omega_n$ in the second, must lie between $\pm\,\omega_a$; all other frequencies are eliminated by the low pass filter.

From formula (14*b*) and the Fourier integral energy theorem, the *noise power* P_N is given by

$$P_N = \frac{N^2}{4\pi^3 T}\int_0^\infty F_p^2 d\omega \int_{\omega-\omega_a}^{\omega+\omega_a}\left(\omega_0 + \omega_n + \frac{\mu}{\lambda}\omega\right)^2 d\omega_n$$
$$+ \frac{N^2}{4\pi^3 T}\int_0^\infty F_m^2 d\omega \int_{-(\omega+\omega_a)}^{-(\omega-\omega_a)}\left(\omega_0 + \omega_n - \frac{\mu}{\lambda}\omega\right)^2 d\omega_n. \tag{17b}$$

Integrating with respect to ω_n, we have

$$P_N = \frac{N^2\omega_a}{2\pi^3 T}\int_0^\infty d\omega\{[(\omega_0 + (1+\nu)\omega)^2 + \tfrac{1}{3}\omega_a^2]F_p^2$$
$$+ [(\omega_0 - (1+\nu)\omega)^2 + \tfrac{1}{3}\omega_a^2]F_m^2\}, \tag{18b}$$

where $\nu = \mu/\lambda$.

Replacing F_p^2 and F_m^2 in (18*b*) by their values as given by (15*b*) and (16*b*), we get

$$P_N = \frac{\omega_a N^2}{\pi^3 T}\int_0^\infty (\omega_0^2 + (1+\nu)^2\omega^2 + \tfrac{1}{3}\omega_a^2)(F_c^2 + F_s^2)d\omega$$
$$+ 4\frac{\omega_a N^2}{\pi^3 T}\int_0^\infty (1+\nu)\omega_0\omega F_cF_s \cos(\theta_c - \theta_s)d\omega. \tag{19b}$$

To evaluate (19*b*) we make use of the formulas, derived below

$$\frac{1}{\pi T}\int_0^\infty (F_c^2 + F_s^2)d\omega = 1, \tag{20b}$$

$$\frac{1}{\pi T}\int_0^\infty \omega^2(F_c^2 + F_s^2)d\omega = \lambda^2\overline{s^2} = P_S, \tag{21b}$$

$$\frac{1}{\pi T}\int_0^\infty \omega F_cF_s \cos(\theta_c - \theta_s)d\omega \to 0 \text{ as } T \to \infty. \tag{22b}$$

Substitution of (20*b*), (21*b*), (22*b*) in (19*b*) gives for large values of T

$$P_N = (\tfrac{1}{3}\omega_a^2 + \omega_0^2 + (1+\nu)^2\lambda^2\overline{s^2})\omega_a N^2. \tag{23b}$$

Here, for convenience, we have replaced N^2/π^2 of (19*b*) by N^2, so that N^2 of (23*b*) may be defined and regarded as the high frequency *noise power level.*

It remains to establish formulas (20b), (21b) and (22b). From the defining formulas (10b) and (11b) and the Fourier integral energy theorem, we have

$$\begin{aligned} \frac{1}{\pi T}\int_0^\infty F_c^2 d\omega &= \frac{1}{T}\int_0^T \cos^2\left(\lambda\int_0^t s dt\right)dt, \\ \frac{1}{\pi T}\int_0^\infty F_s^2 d\omega &= \frac{1}{T}\int_0^T \sin^2\left(\lambda\int_0^t s dt\right)dt. \end{aligned} \tag{24b}$$

Adding we get (20b).

Now differentiate (10b) and (11b) with respect to t and apply the Fourier integral energy theorem; we get

$$\begin{aligned} \frac{1}{\pi T}\int_0^\infty \omega^2 F_c^2 d\omega &= \frac{1}{T}\int_0^T \lambda^2 s^2 \sin^2\left(\lambda\int_0^t s dt\right)dt, \\ \frac{1}{\pi T}\int_0^\infty \omega^2 F_s^2 d\omega &= \frac{1}{T}\int_0^T \lambda^2 s^2 \cos^2\left(\lambda\int_0^t s dt\right)dt \end{aligned} \tag{25b}$$

and, by addition, we get (21b).

To prove (22b) we note that

$$\begin{aligned} (1+\mu s)\cos\left(\lambda\int_0^t s dt\right) &= \cos\left(\lambda\int_0^t s dt\right) + \frac{\mu}{\lambda}\frac{d}{dt}\sin\left(\lambda\int_0^t s dt\right) \\ &= \frac{1}{\pi}\int_0^\infty \left[F_c\cos(\omega t+\theta_c) + \frac{\mu}{\lambda}\omega F_s\cos(\omega t+\theta_s)\right]d\omega \\ &= \frac{1}{\pi}\int_0^\infty \left[F_c^2 + \left(\frac{\mu}{\lambda}\right)^2\omega^2 F_s^2 \right. \\ &\qquad \left. + 2\frac{\mu}{\lambda}\omega F_c F_s\cos(\theta_c-\theta_s)\right]^{1/2}\cos(\omega t+\Phi)d\omega. \end{aligned} \tag{26b}$$

Consequently, by the Fourier integral energy theorem,

$$\begin{aligned} &\frac{1}{T}\int_0^T (1+\mu s)^2\cos^2\left(\lambda\int_0^t s dt\right)dt \\ &\quad = \frac{1}{\pi T}\int_0^\infty \left[F_c^2 + \left(\frac{\mu}{\lambda}\right)^2\omega^2 F_s^2 + 2\frac{\mu}{\lambda}\omega F_c F_s\cos(\theta_c-\theta_s)\right]d\omega \end{aligned} \tag{27b}$$

and

$$\begin{aligned} &\frac{1}{T}\int_0^T \mu s\cdot\cos^2\left(\lambda\int_0^t s dt\right)dt \\ &\qquad = \frac{1}{\pi T}\left(\frac{\mu}{\lambda}\right)\int_0^\infty \omega F_c F_s\cos(\theta_c-\theta_s)d\omega. \end{aligned} \tag{28b}$$

By simple transformations (28*b*) becomes

$$\begin{aligned}\frac{1}{\pi T}\int_0^\infty \omega F_c F_s \cos\,(\theta_c - \theta_s)d\omega\\ &= \frac{1}{2T}\int_0^T \lambda s dt + \frac{1}{4T}\int_0^T \frac{d}{dt}\sin\left(2\lambda\int_0^t s dt\right)dt\\ &= \frac{1}{2}\lambda\bar{s} + \frac{1}{4T}\sin\left(2\lambda\int_0^T s dt\right)\\ &\to 0 \text{ as } T \to \infty,\end{aligned} \tag{29b}$$

since by hypothesis $\bar{s} = 0$.

We note for reference that

$$-\frac{1}{\pi T}\int_0^\infty F_c F_s \sin\,(\theta_c - \theta_s)d\omega = \frac{1}{2T}\int_0^T \sin\left(2\lambda\int_0^t s dt\right)dt. \tag{30b}$$

Bibliography

1. Carson, J. R., "Notes on the Theory of Modulation," *Proc. I. R. E.*, **10**, pp. 57–64, Feb., 1922.
2. Roder, H., "Über Frequenzmodulation," *Telefunken-Zeitung*, **10**, pp. 48–54, Dec., 1929.
3. Heilmann, A., "Einige Betrachtungen zum Problem der Frequenzmodulation," *Elek. Nach. Tech.*, **7**, pp. 217–225, June, 1930.
4. Van der Pol, B., "Frequency Modulation," *Proc. I. R. E.*, **18**, pp. 1194–1205, July, 1930.
5. Eckersley, T. L., "Frequency Modulation and Distortion," *Exp. Wireless and Wireless Engg.*, **7**, pp. 482–484, Sept., 1930.
6. Runge, W., "Untersuchungen an amplituden- und frequenz-modulierten Sendern," *Elek. Nach. Tech.*, **7**, pp. 488–494, Dec., 1930.
7. Roder, H., "Amplitude, Phase and Frequency-Modulation," *Proc. I. R. E.*, **19**, pp. 2145–2175, Dec., 1931.
8. Andrew, V. J., "The Reception of Frequency Modulated Radio Signals," *Proc. I. R. E.*, **20**, pp. 835–840, May, 1932.
9. Barrow, W. L., "Frequency Modulation and the Effects of a Periodic Capacity Variation in a Non-dissipative Oscillatory Circuit," *Proc. I. R. E.*, **21**, pp. 1182–1202, Aug., 1933.
10. Barrow, W. L., "On the Oscillations of a Circuit Having a Periodically Varying Capacitance; Contribution to the Theory of Nonlinear Circuits with Large Applied Voltages," *Proc. I. R. E.*, **22**, pp. 201–212, Feb., 1934, also *M. I. T. Serial* 97, Oct., 1934.
11. Chaffee, J. G., "The Detection of Frequency-Modulated Waves," *Proc. I. R. E.*, **23**, pp. 517–540, May, 1935.
12. Armstrong, E. H., "A Method of Reducing Disturbances in Radio Signaling by a System of Frequency-Modulation," *Proc. I. R. E.*, **24**, pp. 689–740, May, 1936.
13. Crosby, M. G., "Frequency Modulation Propagation Characteristics," *Proc. I. R. E.*, **24**, pp. 898–913, June, 1936.
14. Crosby, M. G., "Frequency-Modulated Noise Characteristics," *Proc. I. R. E.*, **25**, pp. 472–514, April, 1937.
15. Roder, H., "Noise in Frequency Modulation," *Electronics*, **10**, pp. 22–25, 60, 62, 64, May, 1937.